Hammill Institute on Disabilities

Journal of Positive Behavior Interventions
2022, Vol. 24(2) 156–168

DOI: 10.1177/1098300720946651
jpbi.sagepub.com
SAGE

Treatment Components and Participant Characteristics Associated With Outcomes in Self-Monitoring Interventions

Allison Bruhn, PhD[1], Allison Gilmour, PhD[2], Ashley Rila, MA[1], Alyssa Van Camp, MA[3], Amanda Sheaffer, MA[3], Eleanor Hancock, MA[3], Josephine Fernando, PhD[4], and Joseph Wehby, PhD[3]

Abstract

Self-monitoring is one of the most widely used and widely researched strategies for improving student behavior. However, specific research-based guidance about how to design effective self-monitoring interventions and to whom they should be delivered does not yet exist. To this end, we examined how various treatment components and participant characteristics moderated response to self-monitoring interventions. We included 66 single-case studies on academic engagement and 21 single-case studies on disruptive behavior. These studies included 290 participants with challenging behavior, 183 of whom had a disability. After extracting raw data from original studies, we analyzed data using multilevel modeling for each dependent variable (i.e., academic engagement, disruptive behavior). Across both dependent variables, student age and educational setting impacted treatment effects, as did the inclusion of goal-setting, feedback, and reinforcement. Based on our findings, we describe implications related to designing self-monitoring interventions. We also discuss limitations and future directions.

Keywords

classroom intervention(s), self-management, self-regulation, single-case designs data analysis

Students' academic engagement has been linked to levels of academic performance (Elliott, 2019). As a psychological construct, *academic engagement* refers to students' motivation and interest in participating in educational activities, as well as their pride in and attachment to school (Konold et al., 2018). As a measurable and observable construct, *academic engagement* refers to behaviors demonstrating interaction with curricular content such as students attending to the assigned task, following directions, raising hands, answering questions, and completing work (Cooper & Scott, 2017). School-based intervention research on students with challenging behaviors usually takes a behavioral approach to engagement, particularly when the goal is to examine the observable effects of intervention on behavior. Theoretically, if an intervention can positively impact students' academic engagement, then academic success will follow. This pattern of change has been documented experimentally, with research consistently demonstrating high engagement leads to better academic achievement (Konold et al., 2018). Unfortunately, academic engagement and, in turn, achievement may be a formidable task for students with and without disabilities, especially when they struggle with disruptive behaviors.

In addition to impacting the achievement of students with challenging behavior, disruptive behaviors such as blurting out, talking to peers about nonacademic topics, wandering around the room, making inappropriate noises, arguing with others, and playing with materials can be a distraction to classmates and teachers. Often, the teacher stops teaching to attend to disruptive behavior, which interrupts the flow of student learning and, ultimately, can result in student failure (Gage et al., 2018). Typically, severe

[1]University of Iowa, Iowa City, USA
[2]Temple University, Philadelphia, PA, USA
[3]Vanderbilt University, Nashville, TN, USA
[4]Winona State University, MN, USA

Corresponding Author:
Allison Bruhn, University of Iowa, N252 Lindquist Center, Iowa City, IA 52242, USA.
Email: allison-bruhn@uiowa.edu

Action Editor: Paul Caldarella

disruptive behavior is expressed as early as 5 years of age, with early disruptive behavior problems predicting difficulties later in life, including childhood- and adult-onset mental disorders (Wakschlag et al., 2017). Thus, there is a critical need to provide effective intervention strategies to help reduce the potential long-term detrimental effects of disruptive behavior. Like academic engagement, disruptive behavior may be seen across a range of students, including, but not limited to, those with learning disabilities (LD), emotional/behavioral disorders (EBD), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), speech/language impairments (SLI), intellectual disabilities (ID), and neurotypical development.

Self-Monitoring

High-leverage, evidence-based classroom management strategies such as opportunities to respond, behavior specific feedback, and active supervision should be used to prevent disengagement and disruptive behavior, but students with histories of persistent challenging behavior often need additional classroom interventions or supports (McLeskey et al., 2017). One of the most widely used and researched strategies for reducing individual problem behaviors in the classroom is self-monitoring (Carter et al., 2011). Self-monitoring is a meta-cognitive, antecedent-based strategy that involves teaching students to think about and be aware of a predetermined behavior, evaluate the extent to which that behavior occurs, and then record that evaluation either electronically or on a paper form. For students who struggle with self-regulation, self-monitoring prompts them to be intentional about exercising control over their own behavior, thus resulting in improved behavior and productivity (Bandura, 1991).

Numerous systematic literature reviews on self-monitoring and related interventions (e.g., self-evaluation, goal-setting) have presented narrative and descriptive findings related to study characteristics (e.g., participants, setting, experimental design, treatment integrity; Bruhn et al., 2015), treatment outcomes (e.g., reading skills, math skills, behavior; Mooney et al., 2005), and self-monitoring intervention components (e.g., prompts, reinforcement, goal-setting, feedback; Bruhn et al., 2015; Sheffield & Waller, 2010). Across syntheses, researchers have found sizable effects on math (Mooney et al., 2005) and reading skills (Briesch & Chafouleas, 2009) with students who had a range of disabilities (e.g., EBD, ADHD, SLI, LD). Recent reviews also reported overall positive effects on behaviors such as on-task, disruption, and social interactions (Bruhn et al., 2015; Sheffield & Waller, 2010).

As noted in previous self-monitoring reviews, self-monitoring is rarely implemented as a “stand-alone” intervention but rather as a multicomponent package. In a review of single-case design studies of self-monitoring to reduce problem behavior, Sheffield and Waller (2010) found self-monitoring often was paired with reinforcement, with students being rewarded for goal attainment or accurate self-monitoring (e.g., matching teachers’ recording of the same behaviors). In a similar review of self-monitoring studies, adult feedback was provided to students in 25 of 41 studies and was related to students’ accuracy and goal attainment (Bruhn et al., 2015). These external contingencies are generally teacher-managed, which may limit the degree to which the self-monitoring intervention is truly self-managed by the student (Briesch & Chafouleas, 2009). Yet, they may be helpful in promoting improved behavior, self-monitoring accuracy, and generalization across settings (Peterson et al., 2006).

Although these reviews provide valuable information to researchers and practitioners about potentially important components to include when designing a self-monitoring intervention, it is unclear the extent to which various components are related to the effectiveness of the intervention. For example, results of previous reviews indicated that feedback and reinforcement related to students’ goals often were included in self-monitoring interventions. Yet, it is unknown if and how goal-setting improves response to self-monitoring interventions, or whether feedback and reinforcement differentially influence the effectiveness of self-monitoring. In a recent study on data-based decision making about student responsiveness in self-monitoring interventions, using multilevel modeling across 13 single-case designs, authors found increasing goals over time resulted in marginal (0.06% points) improvement in behavior (Bruhn et al., 2020). It is possible these findings can be attributed to ceiling effects such that when students reached a certain level of positive behavior, further growth could not be achieved. Educational and psychological researchers have described the importance of setting goals and receiving feedback toward goal attainment (Covington, 2000; Locke & Latham, 2002), and goals are often included as part of multicomponent self-monitoring interventions. Analyzing the moderating effects of goals, feedback, and reinforcement on behavioral outcomes may help deduce whether these components are actually necessary when designing effective self-monitoring interventions. Furthermore, these data can provide researchers and practitioners relevant information to aid in adapting interventions to enhance student success.

Two other aspects of designing a self-monitoring intervention are determining *interval length* (i.e., the interval of time between instances of students recording their behavior) and *session length* (i.e., how long the self-monitoring session lasts). In Bruhn and colleagues’ (2020) examination of the effects of self-monitoring intervention adaptations, they found increasing interval lengths slowly over time actually worsened behavior, albeit not significantly. Currently, no research-based recommendations for effective interval lengths exist. Researchers and practitioners, in turn, have little guidance for designing effective self-monitoring interventions that include appropriate interval and session lengths. Nor do they have guidance about how to adapt interval lengths based on responsiveness or classroom context.

Although research on self-monitoring has generally demonstrated positive effects for a range of students, the degree to which student characteristics such as gender, age, race, and disability moderate response to intervention has yet to be examined. Existing evidence of disproportionate exclusionary discipline practices for male students, students of color (e.g., Black, Hispanic, American Indian), and students with disabilities (U.S. Department of Education Office for Civil Rights, 2014), coupled with the correlation between exclusionary discipline (e.g., suspension) and poor student outcomes (e.g., dropout, low achievement), underscores the need for positive and effective interventions that may promote school success. Thus, understanding how various students respond to self-monitoring, and in turn, how to design interventions to increase the likelihood of a positive response is critical.

Purpose

The purpose of this review was to examine whether treatment components and student characteristics moderated the effects of self-monitoring on students' academic engagement and disruptive behavior in single-case design studies of self-monitoring. Self-monitoring studies traditionally have relied on single-case methodology for evaluating effects on behavior (Bruhn et al., 2015; Sheffield & Waller, 2010). Single-case studies may be used to establish evidence-based practices (Kratochwill et al., 2010) and thus serve as the primary design evaluated in this review. Additionally, recent software developments allow for researchers to extract raw data from hard copies of published manuscripts, permitting them to analyze data across a body of literature in more complex ways. In this review, we used multilevel modeling to answer the following research question: To what extent do treatment components (e.g., goal, interval length, session length, feedback, reinforcement) and student characteristics (e.g., gender, age, race, disability, setting) moderate participants' response to self-monitoring interventions?

Method

Article Selection Procedures

Search procedures. We used procedures from previous reviews (Bruhn et al., 2015; Van Camp et al., 2020), which have relied on Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards, as a guide for conducting article searches, selection, and coding. First, we conducted an extensive electronic search using ERIC (Education Resources Information Center) and PsycInfo to identify peer-reviewed articles from 2012 to 2019 using a combination of the following search terms: *self-monitoring, self-recording, self-evaluation, self-regulation, self-reporting, and behavior*. Because a recent literature review on self-monitoring (Bruhn et al., 2015) included articles published from 2000 to 2012, we selected articles published during the years 2012–2019. We updated the search beginning in 2012 to ensure no articles were missed in that year. In addition, we did an ancestral search of the articles included in the Bruhn et al. (2015) review; therefore, the current review spans 2000–2019. We obtained a total of 3,949 articles through the electronic search and the Bruhn et al. (2015) literature review. To reduce the potential for publication bias, we also searched ProQuest Dissertations and Theses using the same search terms to identify unpublished literature for inclusion. Specifically, we searched unpublished studies due to indications that published studies in education and special education are associated with significantly higher effect sizes than unpublished studies (d = 0.64; Chow & Ekholm, 2018). This search yielded 5,602 articles. Next, we examined all article titles across the published and unpublished literature and deleted any duplicates. If the title or the abstract indicated a self-monitoring intervention was implemented and behavioral outcomes were measured, we retained the article and two members of the research team read it in its entirety to determine if the study met the inclusion criteria. If there were discrepancies about whether to include the article, the first and last authors read the article and discussed with the group until reaching consensus (100% reliability). Of the 239 articles read in entirety, 66 met inclusion criteria for academic engagement and 21 met criteria for disruptive behavior as the dependent variable (DV), respectively. A list of all included articles is available from the first author.

Inclusion and exclusion criteria. To be included, the article had to use a single-case research design. Outcomes had to be reported in a line graph depicting data from each session and at the student level to allow for raw data extraction. Due to our data analytic methods, we excluded other study designs (e.g., experimental and quasi-experimental group designs, case studies, qualitative studies, literature reviews, meta-analyses). The single-case design had to include one A (i.e., baseline) and one B (i.e., self-monitoring intervention) phase adjacent to each other. We excluded studies without an AB combination (e.g., alternating treatment design) that did not include an initial baseline phase.

Second, the independent variable (IV) had to be a self-monitoring intervention defined as participants thinking about and then recording the extent to which they displayed the behavior(s). We excluded studies in which students self-monitored an academic skill (e.g., number of words or digits correct) rather than a behavioral skill (e.g., on-task). Intervention terminologies could vary. For example, instead of self-monitoring, it could have been described as self-recording (e.g., Moore et al., 2010), as long as the participant observed and recorded the presence or absence of their target behavior(s). If self-monitoring was one component of a multicomponent intervention, the study was included. For example, in addition to self-monitoring, some studies included components such as reinforcement (e.g., Barry & Messer, 2003) to enhance self-monitoring procedures.

Third, the DV had to be a measure of academic engagement or disruptive behavior graphed at the individual student level (not classroom or group level). We excluded studies that included data graphed at the classroom or group level. Studies could include the terms *academic engagement* or *on-task* (e.g., Battaglia et al., 2015) to indicate behavior demonstrated by attending to the assigned task. If the DV was described as disruptive behavior with examples such as talking out of turn, making inappropriate noises, or wandering around the room and distracting others, the study was eligible for inclusion. In addition, we included studies with multiple DVs when one of the outcome variables was a measure of academic engagement or disruptive behavior, and the data were disaggregated by DV. Academic engagement and disruptive behavior had to be reported as a percentage to allow for a common metric across studies. The DV could have been measured using interval recording, duration, or event recording, so long as a percentage was reported and graphed (i.e., in a line graph).

Fourth, participants had to manifest behavioral problems. For example, the participant could have (a) been nominated for displaying problem behaviors (e.g., Amato-Zech et al., 2006); (b) had a diagnosis such as ADHD or other EBD (e.g., Barry & Messer, 2003); or (c) had a behavioral screening score indicating risk (e.g., A. Bruhn & Watt, 2012). Receiving special education services was not a criterion for inclusion; thus, the sample of participants represents students with a range of abilities and behaviors in different educational settings.

Fifth, the study had to take place in an education setting, not a clinic. This included general education (e.g., Wills & Mason, 2014) and special education (e.g., Clemons et al., 2016) classrooms and alternative or residential educational facilities (e.g., Axelrod et al., 2009).

Finally, to provide an update to previous reviews (e.g., Bruhn et al., 2015) the articles had to be published between 2000 and 2019. Sixty-six articles met these inclusion criteria for academic engagement (40 peer-reviewed articles and 26 dissertations/theses) and 21 articles for disruptive behavior (11 peer-reviewed and 10 dissertations/theses). After reaching 100% agreement on inclusion, we subsequently coded included articles using the following procedures.

Coding and Data Extraction Procedures

We coded articles on 25 different variables across nine categories: demographic information (i.e., age, gender, grade level, race/ethnicity, disability status), intervention setting (e.g., general education classroom, special education classroom), treatment fidelity (e.g., reported, percentage), DV reliability (e.g., reported, percentage), feedback (e.g., visual or graphic, praise, correction, encouragement), reinforcement (e.g., any specified component of the intervention that included access to tangibles, attention beyond praise, sensory experience, or task escape), intervention session length (i.e., duration the student self-monitored), interval length (i.e., time between self-monitoring occurrences within the same session), and goals (i.e., presence or absence of a goal, and the specific goal percentage if reported). We developed coding forms with variable definitions. After reviewing and discussing the forms, each coder (i.e., two doctoral students and one professor) coded five articles independently and then met to discuss their findings. Once they were 100% reliable on these first five articles, they independently coded all remaining articles (i.e., 100% of articles were double-coded). Initial intercoder reliability for variables coded by participant was 97.13% (i.e., [16 variables $\times$ 290 participants = 4,640] − 133 discrepancies/4,640 = .9713). Initial intercoder reliability for variables coded by study was 98.21% (i.e., [9 variables $\times$ 87 studies] − 14 discrepancies/783 = .9821). We discussed all disagreements until 100% agreement was achieved for every variable of every included article.

Data extraction. In addition to coding study variables, coders used Plot Digitizer (2015) to extract data from single-case graphs. Plot digitizing software requires researchers to import graphs, calibrate X and Y axes, and click each data point in a data series (Drevon et al., 2017; Shadish et al., 2009). The software then outputs all of the Y values in that data series. A recent study assessing the reliability of plot digitizing software found that 92% of data points digitized by primary and secondary coders were within 1% of agreement (Drevon et al., 2017). In this review, a primary coder and secondary coder extracted data. Coders digitized only the first A and first B condition of the graphs for use in the analysis. Coders corrected obvious errors that occurred in the extraction process in the following ways: (a) negative values on outcomes in which the minimum Y value was 0 were corrected to 0 and (b) values that exceeded the maximum Y value were corrected to the maximum value (Zimmerman et al., 2018).

Coders employed point-by-point agreement to calculate intercoder reliability (i.e., total number of agreements over total number of agreements plus disagreements multiplied by 100). When primary and secondary coders' values differed by more than 2% of the maximum value of the range on the Y axis of a graph, the graphs were digitized by a third coder. Initial coding agreement between the first two coders was 87.8% for academic engagement and 91.2% for disruptive behavior. A third coder extracted data for any data points falling outside the 2% window of agreement. One hundred percent of the third coder's extractions were within the 2% window of agreement with the original coder.

Quantitative Analyses

We evaluated whether treatment components or participant characteristics moderated the treatment effects of self-monitoring interventions using multilevel models. Multilevel modeling accounts for the nesting of data within students and nesting of students in studies and allows for researchers to examine the influence of variables on treatment effects in single-case designs (Moeyaert et al., 2014). This approach also allowed us to include slope parameters to account for data trends at baseline and intervention. We used models with the percentage of time that a participant was academically engaged and the percentage of time that a participant exhibited disruptive behavior as the DVs. Researchers have previously used this type of modeling when synthesizing single-case designs in behavioral research (e.g., Bruhn et al., 2020; Gage et al., 2012). The results are interpreted as the percentage point change in academic engagement or disruptive behavior due to the predictor of interest, after accounting for trends in the data. We fit all models in Stata 15 using restricted maximum likelihood and calculated degrees of freedom using the Kenward–Roger method. The Kenward–Roger method is appropriate for small samples and recommended when using multilevel models to synthesize single-case designs (Ferron et al., 2009). In addition, we calculated log response ratios to assess the overall effect sizes of self-monitoring on both DVs (Pustejovsky, 2018; Pustejovsky & Swan, 2018).

We fit a series of separate models to examine how each treatment component and participant characteristic was associated with treatment effects. The treatment components we examined included session length in minutes, interval length in minutes, use of goal setting, the percentage point goal when the study included a goal-setting component, use of reinforcement, use of feedback, and use of both reinforcement and feedback. We grand mean centered the session length, interval length, and percentage point goal. We examined treatment moderation by adding an interaction between treatment and treatment component or participant characteristic to the multilevel models. We included an indicator for if the study included only reinforcement, an indicator for if the study included only feedback, and an indicator for if the study included reinforcement and feedback used jointly. Studies that did not include reinforcement or feedback acted as the comparison in this model.

The student characteristics included disability status, gender, race, age, and treatment setting (e.g., general education classrooms or special education setting). We grand mean centered age to have an interpretable intercept and interaction estimate; all other variables were coded as dummy variables with White female participants without disabilities who received treatment in general education classrooms and were the same average age (10.7 years old for academic engagement and 10.2 years for disruptive behavior) as the comparison group.

In each model, we included session number and an interaction between session number and treatment. This approach allowed us to assess the average slope at baseline and the change in slope due to treatment. Our moderator models examined the initial change in level of behavior rather than the change in behavior over time due to treatment. We chose to examine the association between the moderators and the average change in level because self-monitoring was expected to have an immediate effect on academic engagement and disruptive behavior given these behaviors are generally reversible and related to performance, rather than skill, deficits. We included only the first AB phase in the quantitative analyses.

Results

The results of the quantitative analyses are presented in Tables 1–4. The sample sizes change in each model because some studies did not report information regarding all of the moderators. One study included data from the same participants in multiple settings (Bruhn et al., 2017). To avoid introducing cross-classification into the models, we randomly selected a single setting for each participant in that study and only included data for the participant in that setting in the analyses. Across all models, the intraclass correlations (ICCs) and random effects supported our use of a three-level model. We calculated the ICC from an empty model and found 13.2% of the total variation in academic engagement was between studies and 8.3% of the variation in academic engagement was between participants. In the empty model we fit for disruptive behavior, 21.5% of the total variation in disruptive behavior was between studies and 15.7% was between participants. The random effects in all models were substantively large, suggesting substantial variation between studies and participants even after accounting for treatment effects, slopes at baseline and treatment, and moderators.

Academic Engagement

To contextualize findings, we coded for DV reliability and treatment fidelity. In 61 of 66 studies, reliability of academic engagement was reported (e.g., interobserver agreement [IOA] or Kappa). Across studies, mean IOA was

Table 1. Multilevel Models Examining Academic Engagement Treatment Component Moderations.

Components	No predictors	Session length	Interval length	Goal	Goal percentage	Reinforcement + feedback
Treatment (Tx)	33.54*** (0.89)	35.62*** (0.97)	33.76*** (0.95)	32.50*** (0.92)	33.79*** (1.72)	30.48*** (1.23)
Session length		0.01 (0.01)				
Session × Tx		−0.01 (0.01)				
Interval length			0.11 (0.10)			
Interval × Tx			−0.11* (0.04)			
Goal				−1.27 (2.97)		
Goal × Tx				5.76*** (1.27)		
Goal%					0.01 (0.15)	
Goal% × Tx					0.53*** (0.09)	
Reinforcement						−0.90 (4.49)
Reinforcement × Tx						8.93*** (1.70)
Feedback						−0.83 (3.80)
Feedback × Tx						6.42*** (1.55)
Reinforcement + feedback						3.60 (3.16)
Reinforcement + FB × Tx						2.68 (1.43)
Session	0.18* (0.07)	0.17* (0.08)	0.18* (0.07)	0.19** (0.07)	−0.17 (0.14)	0.23** (0.07)
Session × Tx	0.04 (0.08)	0.02 (0.08)	0.04 (0.08)	0.01 (0.08	0.39** (0.14)	−0.05 (0.08)
Intercept	45.25*** (1.36)	44.37*** (1.47)	46.01*** (1.40)	45.45*** (1.57)	47.19*** (2.06)	44.01*** (2.40)
Study RE	78.79 (19.05)	72.94 (20.68)	72.59 (19.40)	79.95 (19.50)	37.88 (20.80)	77.83 (19.43)
Student RE	61.07 (8.68)	63.22 (9.99)	60.86 (9.13)	61.35 (8.72)	37.69 (13.01)	61.25 (8.69)
Residual	287.15 (6.17)	289.72 (6.89)	286.60 (6.54)	285.82 (6.14)	290.41 (12.67)	285.15 (6.13)
Studies	66	54	58	66	18	66
Participants	225	183	202	225	53	225
Obs.	4,555	3,726	4,049	4,555	1,109	4,555

Note. Standard errors are in parentheses. Significance test reflects degrees of freedom adjusted using the Kenward–Roger method. FB = feedback; RE = random effects; Tx = treatment.
*$p < .05$. **$p < .01$. ***$p < .001$.

94.60% (SD = 3.23) and mean Kappa was .84 (SD = .06). Treatment fidelity of the self-monitoring interventions was reported in 39 of 66 studies, with a mean of 96.26% (SD = .07) and 37 of 39 reporting average fidelity greater than 85%.

Treatment components. Self-monitoring intervention sessions ranged from 5 to 420 min (M = 54.26, SD = 94.70, Mdn = 27.50), though the majority of sessions were under 60 min. The interval length (i.e., the time between self-monitoring instances) ranged from every 30 s to every 84 min (M = 6.60, SD = 12.21, Mdn = 3.00). Sixty participants did not receive feedback or reinforcement, 36 received reinforcement only, 51 received feedback only, and 79 received both feedback and reinforcement. Of the 85 participants who had a goal tied to their self-monitoring intervention, those goals ranged from 20% to 88% (e.g., earn 80% of points on self-monitoring form; M = 75.15, SD = 12.78).

Participant characteristics. Across the 66 studies in which the DV was academic engagement, there were 228 student participants; 173 males and 55 females. For 173 participants, ages ranged from 6 to 17 years (M = 10.65, SD = 3.13), with age not reported for 55 participants. Race/ethnicity was not reported for 81 participants, and 82 were reported as White, 34 as Black, 16 as Hispanic, three as Asian, and 12 as Other. Participants included 148 with an identified disability or disorder (e.g., LD, ADHD, ASD, ID, SLI, EBD), with 117 of these students receiving special education services. Authors reported grade level for 203 participants, with grades ranging from pre-kindergarten (preK) to 12th grade. Four participants were in preK schools, 130 were in elementary school, 40 were in middle school, and 24 were in high school (30 = not reported). The majority of participants received self-monitoring interventions in general education settings (n = 148), whereas 75 were special education classrooms, and five were not reported.

Multilevel models. The first column of Table 1 presents the results of a three-level multilevel model without any hypothesized moderators in the model; we included session number and an interaction between session number in

Table 2. Multilevel Models Examining Academic Engagement Participant Characteristic Moderations.

Components	Disability status	Gender	Race	Age	Setting
Treatment (Tx)	32.18*** (1.14)	29.69*** (1.30)	33.53*** (1.36)	34.16*** (1.20)	36.12*** (0.99)
Disability	−5.02* (2.07)				
Disability × Tx	2.15 (1.14)				
Male		−4.73** (1.80)			
Male × Tx		5.12*** (1.25)			
Black			1.21 (3.02)		
Hispanic			−1.22 (3.37)		
Other			−1.43 (3.78)		
Black × Tx			−1.12 (1.72)		
Hispanic × Tx			1.00 (2.33)		
Other × Tx			1.98 (2.44)		
Age				1.31* (0.50)	
Age × Tx				−0.83*** (0.22)	
Special education					5.88* (2.40)
Special education × Tx					−5.19*** (1.15)
Session	0.19** (0.07)	0.18* (0.07)	0.25** (0.09)	0.40*** (0.09)	0.28*** (0.08)
Session × Tx	0.03 (0.08)	0.04 (0.08)	−0.05 (0.10)	−0.08 (0.10)	−0.12 (0.08)
Intercept	48.57*** (1.95)	48.86*** (1.93)	44.93*** (2.09)	42.53*** (1.56)	42.79*** (1.64)
Study RE	83.50 (20.50)	78.53 (19.02)	93.00 (28.82)	70.95 (22.07)	81.65 (19.67)
Student RE	58.80 (8.49)	61.28 (8.71)	70.51 (12.44)	71.79 (11.85)	60.03 (8.61)
Residual	286.96 (6.17)	286.10 (6.15)	312.76 (8.31)	317.73 (8.20)	279.07 (6.11)
Studies	66	66	44	50	65
Students	225	225	147	167	220
Obs.	4,555	4,555	2,814	3,167	4,397

Note. Standard errors are in parentheses. Significance test reflects degrees of freedom adjusted using the Kenward–Roger method. RE = random effects; Tx = treatment.
*$p < .05$. **$p < .01$. ***$p < .001$.

treatment to account for slopes at baseline and intervention. Participants were, on average, academically engaged for 45.25% of intervals at baseline. On average, each session at baseline was associated with a 0.19% point change in academic engagement ($p < .01$), an average slope that did not change at treatment. Academic engagement increased by 33.54% points ($LRR_i = 0.62$; $p < .001$) to an average of 78.79% of intervals at treatment, after accounting for slopes at baseline and intervention.

The remaining columns of Table 1 present the results of moderator analyses related to treatment components; all models account for slopes at baseline and intervention. Session length was not associated with different levels of academic engagement at treatment. In contrast, studies that included longer intervals had smaller average treatment effects with a 1-min increase in interval length associated with a −0.11% point change in academic engagement at treatment ($p < .05$). Participants in studies that included goal-setting had treatment effects that were, on average, 5.76% points higher than participants in studies that did not include goal-setting ($p < .001$). Among studies with goal setting, 18 studies with 53 participants, a higher goal was associated with a larger treatment effect with a 1% point increase in the goal associated with a 0.53% point increase in engagement at treatment ($p < .001$). When studies included reinforcement as part of the intervention, the average treatment effects were 8.93% points higher than in studies that did not include reinforcement or feedback ($p < .001$). Participants in studies that included feedback as part of the intervention had higher average academic engagement at treatment than studies that did not include reinforcement or feedback ($b = 6.42$, $p < .001$). There did not appear to be a benefit of adding both reinforcement and feedback to the intervention.

The columns of Table 2 present the results of moderator analyses related to participant characteristics accounting for slopes during baseline and intervention. Participants with disabilities had lower average academic engagement at baseline compared with participants without disabilities. Average treatment effects were the same across participants with and without disabilities. Male participants had lower average academic engagement than female participants at baseline, but male participants had higher average treatment effects ($b = 5.12$, $p < .001$). Participants did not vary by race/ethnicity in average academic engagement at baseline and treatment effects were similar across all races/

Table 3. Multilevel Models Examining Disruptive Behavior Treatment Component Moderations.

Components	No predictors	Session length	Interval length	Goal	Goal percentage	Reinforcement + feedback
Treatment (Tx)	−29.00*** (1.72)	−26.17*** (1.92)	−30.99*** (1.87)	−32.58*** (1.86)	−20.31*** (2.67)	−14.69** (2.85)
Session length		−0.00 (0.02)				
Session × Tx		−0.01 (0.01)				
Interval length			−0.07 (0.09)			
Interval × Tx			0.00 (0.03)			
Goal				−13.00* (5.65)		
Goal × Tx				10.11*** (2.04)		
Goal%					−0.00 (0.00)	
Goal% × Tx					0.002* (0.00)	
Reinforcement						7.30 (13.31)
Reinforcement × Tx						−25.44*** (3.84)
Feedback						−7.89 (10.64)
Feedback × Tx						−16.73*** (3.58)
Reinforcement + feedback						3.08 (7.44)
Reinforcement + FB × Tx						−17.93*** (2.95)
Session	−0.42*** (0.11)	−0.34** (0.12)	−0.47*** (0.12)	−0.44*** (0.11)	−0.45* (0.19)	−0.56*** (0.12)
Session × Tx	0.77*** (0.13)	0.63*** (0.14)	0.87*** (0.14)	0.81*** (0.13)	0.62** (0.22)	1.00*** (0.14)
Intercept	39.37*** (2.85)	36.35*** (2.98)	40.81*** (3.27)	43.69*** (3.33)	31.86*** (436)	37.85*** (6.60)
Study RE	107.55 (48.28)	81.47 (49.01)	123.88 (61.15)	100.57 (45.73)	49.85 (93.75)	106.90 (53.71)
Student RE	105.99 (26.54)	124.14 (34.14)	122.19 (32.70)	103.17 (25.70)	192.62 (80.31)	108.86 (27.19)
Residual	253.37 (10.53)	264.97 (12.09)	262.06 (11.61)	248.52 (10.33)	180.60 (13.48)	243.49 (10.14)
Studies	21	17	18	21	7	21
Students	62	50	53	62	21	62
Obs.	1,223	1,016	1,076	1,223	384	1,223

Note. Standard errors are in parentheses. Significance test reflects degrees of freedom adjusted using the Kenward–Roger method. FB = feedback; RE = random effects; Tx= treatment.
*$p < .05$. **$p < .01$. ***$p < .001$.

ethnicities. A 1-year increase in age was associated with an increase in the percentage of intervals that participants were academically engaged at baseline, but an average decrease in the treatment effect of 0.83% of intervals ($p < .001$). Participants in special education classrooms (e.g., resource rooms, self-contained classrooms, or residential settings) were more engaged than participants in general education settings at baseline. Participants in special education settings experienced lower average treatment effects than participants in general education classrooms ($b = -5.19$, $p < .001$).

Disruptive Behavior

In 20 of 21 studies, reliability of disruptive behavior was reported. Mean IOA was 93.75% (SD = 4.39) and mean Kappa was .83 (SD = .08). Treatment fidelity of the self-monitoring interventions was reported in 10 of 21 studies, with a mean of 93.69% (SD = .07) and 9 of 10 reporting average fidelity greater than 85%.

Treatment components. Self-monitoring intervention sessions ranged from 10 to 420 min (M = 133.06, SD = 175.07, Mdn = 37.5); the majority of sessions were under 60 min. The interval length ranged from every 30 s to every 140 min (M = 17.97, SD = 32.31, Mdn = 5.00); the majority were 10 min or less. Nine participants received only feedback, five received only reinforcement, 38 received both feedback and reinforcement, and 10 received neither. Of the 37 students who had a goal tied to their self-monitoring intervention, those goals ranged from 12.5% to 88% (e.g., earn 80% of points on self-monitoring form; M = 57.92, SD = 29.27).

Participant characteristics. Across 21 studies in which the DV was disruptive behavior, there were 62 student participants; 46 males and 16 females. For 50 participants, ages ranged from 5 to 17 years (M = 10.02, SD = 2.75), with age not reported for 12 participants. Authors did not report race/ethnicity for 17 participants, and 24 were reported as White, 17 as Black, two as Hispanic, one as Asian, and one

Table 4. Multilevel Models Examining Disruptive Behavior Participant Characteristic Moderations.

Components	Disability status	Gender	Race	Age	Setting
Treatment (Tx)	−27.56*** (2.36)	−26.46*** (2.36)	−17.61*** (2.09)	−27.91*** (1.91)	−30.67*** (2.03)
Disability	9.21* (4.47)				
Disability × Tx	−1.92 (2.14)				
Male		4.03 (4.04)			
Male × Tx		−3.31 (2.09)			
Black			14.52* (5.76)		
Hispanic			−4.93 (10.58)		
Other			4.05 (14.58)		
Black × Tx			−14.35*** (2.19)		
Hispanic × Tx			2.03 (5.25)		
Other × Tx			−7.27 (8.33)		
Age				−1.13 (1.29)	
Age × Tx				1.19* (0.48)	
Special education					−2.84 (5.78)
Special education × Tx					5.00* (2.39)
Session	−0.43*** (0.11)	−0.41*** (0.11)	−0.38*** (0.12)	−0.46*** (0.12)	−0.45*** (0.12)
Session × Tx	0.77*** (0.13)	0.75*** (0.13)	0.60*** (0.14)	0.80*** (0.14)	0.83*** (0.14)
Intercept	32.70*** (4.26)	36.42*** (4.10)	29.41*** (3.83)	37.66*** (3.32)	40.37*** (3.60)
Study RE	97.90 (45.75)	101.97 (47.46)	66.05 (46.33)	120.48 (59.26)	114.07 (59.72)
Student RE	101.41 (25.98)	109.00 (27.53)	129.19 (37.43)	91.27 (26.07)	118.20 (32.21)
Residual	253.48 (10.55)	253.08 (10.53)	228.98 (10.82)	245.33 (11.36)	267.33 (11.57)
Studies	21	21	14	16	20
Students	62	62	45	50	57
Obs.	1,223	1,223	947	990	1,130

Note. Standard errors are in parentheses. Significance test reflects degrees of freedom adjusted using the Kenward–Roger method. RE = random effects; Tx = treatment.
*$p < .05$. **$p < .01$. ***$p < .001$.

as Other. Thirty-five participants had an identified disability or disorder (e.g., LD, ADHD, EBD, SLI), with 25 students receiving special education services. Authors reported grade levels for 50 participants, with grades ranging from kindergarten to 11th grade. Thirty-nine participants were in elementary school, 15 were in middle school, and three were in high school (5 = not reported). The majority of participants received self-monitoring interventions in general education settings (n = 40), whereas 17 occurred in special education settings, and five were not reported.

Multilevel models. Tables 3 and 4 include the full results of the multilevel models examining disruptive behavior. At baseline, participants exhibited disruptive behavior during an average of 39.37% of intervals. During baseline, each session was associated with an average −0.42% point change in the percentage of intervals with disruptive behavior ($p < .001$); treatment increased the slope by 0.77% points per session ($p < .001$). During treatment, there was an average decrease of 29.00% points in disruptive behavior (LRR_d= 0.34; $p < .001$). On average, students exhibited disruptive behavior during 10.37% of intervals during treatment, after accounting for average slopes at baseline and treatment.

Session length and interval length did not appear to moderate the treatment effects, accounting for slopes during baseline and treatment. Studies that included goal setting had lower average disruptive behavior at baseline than studies that did not include goal setting. Treatment effects were smaller during treatment in studies that included in goal setting than in studies that did not include goal setting (b = 10.11, $p < .001$). Of studies that included goal setting, higher goals were related to smaller declines in disruptive behavior during treatment (b = 0.002, $p < .05$). Notably, only seven studies with 21 participants included goal setting so this finding should be interpreted with caution. Participants in studies that included reinforcement exhibited less disruptive behavior during treatment (b = −25.44, $p < .001$) compared with participants in studies that did not include reinforcement or feedback. Feedback was associated with a 16.73 additional percentage point decrease in disruptive behavior during treatment compared with studies without reinforcement or feedback ($p < .001$). When studies simultaneously included reinforcement and feedback students exhibited, on average, 17.93% points less disruptive behavior during treatment than studies that did not include reinforcement or feedback ($p < .001$).

The effects of self-monitoring on disruptive behavior were moderated by race/ethnicity, age, and educational setting, but not disability status or gender, after accounting for slopes during baseline and treatment. These results are reported in Table 4. Students with disabilities had higher levels of disruptive behavior at baseline ($b = 9.21$, $p < .001$) but experienced effects similar to students without disabilities during treatment. Male and female participants had similar levels of disruptive behavior during baseline and treatment. Black participants exhibited more disruptive behavior at baseline ($b = 14.52$, $p < .05$), but had greater reductions in disruptive behavior during treatment compared with White participants ($b = -14.35$, $p < .001$). Each additional year in age was associated with a 1.19% point increase in disruptive behavior during treatment ($p < .05$). Participants in special education and general education settings exhibited similar levels of disruptive behavior at baseline, but students in special education settings exhibited, on average, more disruptive behavior during treatment ($b = 5.00$, $p < .05$).

Discussion

Self-monitoring is one of the most widely used and widely researched interventions for students with challenging behavior (Carter et al., 2011). Multiple systematic reviews have documented the effects of self-monitoring in improving academic and behavioral outcomes for students with varying ages, genders, races, and abilities (e.g., Briesch & Chafouleas, 2009; Bruhn et al., 2015; Mooney et al., 2005; Sheffield & Waller, 2010). Despite the breadth and depth of these reviews, a systematic review examining the moderating effects of student and treatment characteristics has yet to be conducted. By better understanding who is likely to benefit from self-monitoring, as well as the components that are most beneficial, researchers and practitioners may be able to design more effective interventions and make adaptations to existing interventions to increase the likelihood of success. To this end, we discuss findings and implications related to the moderating effects of treatment components and student characteristics on the academic engagement and disruptive behavior of participants receiving self-monitoring interventions. We conclude with limitations of the review and future directions.

Findings and Implications

Overall, self-monitoring interventions produced substantial increases in academic engagement and decreases in disruptive behavior. In practical terms, the 33.54% point increase in engagement equates to an additional 3.4 min of engagement for every 10 min of class the student is participating in self-monitoring. For disruptive behavior, a 29.00% point decrease in behavior equates to nearly 3 fewer min of disruption for every 10 min of class. The implications for teachers are clear, especially as these minutes are multiplied across sessions and days. For instance, in five 10-min self-monitoring sessions a week, we might expect a student to go from 20 of 50 min of engagement to 37 of 50 total min of engagement. In light of the relation between engagement and achievement, these findings are noteworthy and indicate self-monitoring is effective in producing an immediate and impactful change in behavior.

Treatment components. Results related to treatment components varied across DVs. Whereas session length (i.e., the amount of time spent self-monitoring) had no moderating effect on either DV, interval lengths did impact academic engagement. Specifically, self-monitoring interventions with shorter interval lengths produced higher rates of engagement during treatment. It should be noted that interval lengths were treated as a continuous (rather than categorical) variable, and the range of interval lengths was quite large. Based on these findings, we tentatively recommend interval lengths of 5 min or less, so long as this fits within the context of classroom instruction.

Perhaps the most interesting finding in our analysis relates to the moderating effects of goals. Goal-setting is often included as part of self-monitoring intervention “packages.” For instance, students may be self-monitoring their on-task behavior and set a goal to be on task for 70% of intervals. As students meet a goal consistently, the goal may be raised to help the student improve behavior or maintain behavior (Bruhn et al., 2020). Researchers have long asserted the importance of goal-setting for improving behavior (Covington, 2000; Locke & Latham, 2002), and yet, our findings indicate including a goal resulted in smaller changes in disruptive behavior compared with self-monitoring interventions without a goal. Conversely, academic engagement was higher in studies that included a goal. One limitation of our analysis is that we did not code for the appropriateness of the goal. It is possible goals were not realistic and thus less likely to be attained. Similarly, we did not code for self-selected versus adult-selected goals. Whereas most goal-setting research involves adult-selected goals for students, self-selected goals lead to better buy-in and better outcomes (Bruhn et al., 2016). This begs the question as to whether multicomponent self-monitoring interventions are more effective when self-directed versus adult-directed, particularly as it relates to the goal-setting component (Briesch & Chafouleas, 2009).

In addition, feedback and reinforcement often are included in multicomponent self-monitoring interventions (Bruhn et al., 2015; Sheffield & Waller, 2010). Feedback (e.g., correction, instruction, praise, graphic) may be delivered by a peer or an adult and relates to students’ behavior or progress in the intervention. Reinforcement (e.g., attention, token, toy, break, food) may be delivered for accurate self-monitoring, meeting a goal, or following the self-monitoring procedures. Some single-case design studies have examined the effects of self-monitoring with and without these components (e.g., Ardoin & Martens, 2004; Freeman & Dexter-Mazza, 2004), with results varying by individual participant. In our analysis, studies were classified as including feedback only, reinforcement only, both feedback and reinforcement, or neither. The strongest effects on academic engagement were found in studies including reinforcement only followed by feedback only. Including only one of those components appeared to result in stronger effects than including both components or none of those components. Similarly, the biggest decreases in disruptive behavior occurred in studies that included reinforcement only. This was followed by studies including both reinforcement and feedback, then feedback alone. Studies that included neither component showed the weakest effects. These findings, while not surprising given the long-standing research supporting the use of feedback and reinforcement for changing behavior, do raise questions about the extent to which self-monitoring is truly a self-managed process given the magnitude of effects that are dependent upon external contingencies. If the goal is to help students become independent learners who can self-regulate their own behavior absent external contingencies, then further research is needed to determine how to effectively fade feedback and reinforcement to promote long-lasting cognitive and behavioral change via independent self-regulation.

Participant characteristics. For academic engagement and disruptive behavior, we found students with disabilities performed similarly in treatment to those without disabilities, thus indicating self-monitoring may be effective for a wide range of students. This is particularly important given the varying disabilities included in the sample (e.g., LD, ADHD, ASD, ID, EBD, SLI, developmental delay [DD]), as some self-monitoring reviews have focused exclusively on students with or at risk for EBD (Bruhn et al., 2015; Carter et al., 2011; Mooney et al., 2005). Interestingly, however, when the intervention took place in general education settings as compared with a special education setting, participants tended to experience better academic engagement and less disruptive behavior. These smaller effects in special education settings may be attributed to more positive behaviors during baseline; thus, there was less room for improvement. For general educators, these results demonstrate self-monitoring interventions can be highly effective for improving individual students' challenging behavior even in the context of a larger, more complex setting.

In addition, male participants tended to experience larger treatment effects than females for academic engagement only. The age of the student moderated effects on both DVs, with younger students demonstrating larger treatment effects. Although these findings do not suggest self-monitoring is ineffective for female students or older students, they do indicate that males and younger students are likely to respond more positively. Previous reviews have indicated less research on self-monitoring has occurred with middle and high school students (Bruhn et al., 2015; Carter et al., 2011; Mooney et al., 2005), which is unfortunate given middle and high school teachers expect their students to become more independent, self-determined individuals. Thus, further research with this age group is needed to establish effective strategies that support the development of self-determined behaviors (Carter et al., 2011).

In terms of race/ethnicity, Black students had higher rates of disruptive behavior at baseline and greater decreases in treatment than students identified as White, Hispanic, or Other. Race/ethnicity did not moderate the effects on self-monitoring for academic engagement. However, these results should be interpreted with caution as over a quarter of the sample for disruptive behavior and over a third of the sample for academic engagement did not report race/ethnicity. Despite the limited sample, the findings for non-White students are promising as it relates to preventing disproportionate exclusionary discipline practices. That is, rather than referring a student to the office or suspending a student for patterns of problem behavior, teachers may implement practical, feasible, and effective classroom interventions such as self-monitoring. The logic is that keeping students in the classroom and providing proactive strategies to help improve their behavior will make it less likely they will be suspended, fail courses, and in turn, dropout. Further research is needed to examine the long-term effects of classroom interventions (e.g., self-monitoring) in preventing these negative outcomes.

Limitations and Future Directions

Although this study contributes to the field by indicating how self-monitoring interventions should be designed and who may be most responsive to self-monitoring interventions, several limitations must be considered. First, we restricted our analysis to the first AB condition because (a) comparing adjacent conditions is consistent with single-case visual analysis, and (b) some later conditions in more complex designs (e.g., multitreatment) consisted of interventions other than self-monitoring. As such, we did not conduct visual analysis to determine the presence of a functional relation. Consistent with our quantitative analysis, visual analysis of the AB condition would have allowed us to determine only one demonstration of effect—not a functional relation, which requires three. Thus, our analyses account only for initial, immediate changes in behavior compared with baseline, not overall changes in behavior across multiple phases or the rate of change over time. Future reviews may include visual analysis and additional phases; however, only adjacent phases should be compared per single-case methodology. A second limitation is that we used linear probability models instead of logistic models. Linear probability models can sometimes result in out of range linear combinations of coefficients; however, the results are very interpretable and our DVs were approximately normally distributed. A third limitation is that we used log response ratios as an effect size measure of the overall effect of self-monitoring on both DVs, despite having trends in the data. The presence of trends can result in biased log response ratio estimates (Pustejovsky, 2018). With regard to limitations within studies, several studies did not report various treatment components or participant characteristics. Relatedly, we did not code studies for the extent to which they met rigorous single-case design standards (e.g., What Works Clearinghouse), though we did code for reliability of the DV and treatment fidelity, with fidelity being underreported in studies of disruptive behavior. To allow for accurate interpretation of results, study replication, and moderator analyses across studies, authors should report all relevant participant characteristics and describe the intervention components with replicable precision.

Conclusion

Despite the aforementioned limitations, this study contributes to breadth and depth of literature on self-monitoring by offering a nuanced analysis on how treatment components and participant characteristics moderate response to self-monitoring interventions. In sum, we found self-monitoring produced significant initial increases in academic engagement and decreases in disruptive behavior. Across both DVs, participants in general education settings and who were younger experienced equal and sometimes greater benefits from self-monitoring. The most effective self-monitoring interventions tended to be multicomponent and include reinforcement and/or feedback related to student performance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Funding for this paper was provided by the Institute of Education Sciences (510-14-2540-00000-13607400-6200-000-00000-20-0000).

References

Amato-Zech, N., Hoff, K., & Doepke, K. (2006). Increasing on-task behavior in the classroom: Extension on self-monitoring strategies. *Psychology in Schools*, *43*(2), 211–221. https://doi.org/10.1002/pits.20137

Ardoin, S. P., & Martens, B. K. (2004). Training children to make accurate self-evaluations: Effects on behavior and the quality of self-ratings. *Journal of Behavioral Education*, *13*, 1–23. https://doi.org/10.1023/B:JOBE.0000011257.63085.88

Axelrod, M., Zhe, E., Haugen, K., & Klein, J. (2009). Self-management on on-task homework behavior: A promising strategy for adolescents with attention and behavior problems. *School Psychology Review*, *38*(3), 325–333. https://doi.org/10.1080/02796015.2009.12087817

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*, 248–297.

Barry, L., & Messer, J. (2003). Practical application of self-management for students diagnosed with Attention-Deficit/Hyperactivity Disorder. *Journal of Positive Behavior Interventions*, *5*(4), 238–248. https://doi.org/10.1177/10983007030050040701

Battaglia, A., Radley, K., & Ness, E. (2015). Evaluating the effects of on-task in a box as a class-wide intervention. *Psychology in the Schools*, *52*(8), 743–755. https://doi.org/10.1002/pits.21858

Briesch, A. M., & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988-2008). *School Psychology Quarterly*, *24*(2), 106–118. https://doi.org/10.1037/a0016159

Bruhn, A., McDaniel, S., & Kreigh, C. (2015). Self-monitoring interventions for students with behavior problems: A systematic review of current research. *Behavioral Disorders*, *40*(2), 102–121. https://doi.org/10.17988/BD-13-45.1

Bruhn, A., & Watt, S. (2012). Improving behavior by using multicomponent self-monitoring within a targeted reading intervention. *Behavioral Disorders*, *38*(1), 3–17. https://doi.org/10.1177/019874291203800102

Bruhn, A. L., McDaniel, S. M., Fernando, J., & Troughton, L. (2016). Goal-setting for students with significant behavior problems: A systematic review of the literature. *Behavioral Disorders*, *41*(2), 107–121. https://doi.org/10.17988/0198-7429-41.2.107

Bruhn, A. L., Rila, A., Mahatmya, D., Estrapala, S., & Hendrix, N. (2020). The effects of data-based, individualized interventions for behavior. *Journal of Emotional and Behavioral Disorders*, *28*, 3–6. https://doi.org/10.1177/1063426618806279

Bruhn, A. L., Woods-Groves, S., Fernando, J., Choi, T.*, & Troughton, L. (2017). Evaluating technology-based self-monitoring as a tier 2 intervention across middle school settings. *Behavioral Disorders*, *42*(3) 119–131. https://doi.org/10.1177/0198742917691534

Carter, E. W., Lane, K. L., Crnobori, M. E., Bruhn, A. L., & Oakes, W. P. (2011). Self-determination interventions for students with or at risk for emotional and behavioral disorders: Mapping the knowledge base. *Behavioral Disorders*, *36*(2), 100–116. https://doi.org/10.1177/019874291103600202

Chow, J., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review*, 30, 727–744. https://doi.org/10.1007/s10648-018-9437-7

Clemons, L. L., Mason, B. A., Garrison-Kane, L., & Wills, H. P. (2016). Self-monitoring for high school students with disabilities: A cross-categorical investigation of I-Connect. *Journal of Positive Behavior Interventions*, *18*(3), 145–155. https://doi.org/10.1177/1098300715596134

Cooper, J. T., & Scott, T. M. (2017). The keys to managing instruction and behavior: Considering high probability practices. *Teacher Education and Special Education*, *40*(2), 102–113. https://doi.org/10.1177/0888406417700825

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171–200. https://doi.org/10.1146/annurev.psych.51.1.171

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification*, *41*(2), 323–339. https://doi.org/10.1177/0145445516673998

Elliott, L. (2019). Profiles of academic, cognitive, and behavioral development from kindergarten to third grade. *Learning and Individual Differences*, *70*, 109–120. https://doi.org/10.1016/j.lindif.2019.01.010

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobinoff, G., & Hibbard, S. T. (2009). Making effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, *41*(2), 372–384. https://doi.org/10.3758/BRM.41.2.372

Freeman, K. A., & Dexter-Mazza, E. (2004). Using self-monitoring with an adolescent with disruptive classroom behavior: Preliminary analysis of the role of adult feedback. *Behavior Modification*, *28*(3), 402–419. https://doi.org/10.1177/0145445503258982

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional behavioral assessment-based interventions for students with or at risk for emotional and/or behavioral disorders in school: A hierarchical modeling meta-analysis. *Behavioral Disorders*, *37*(2), 55–77. https://doi.org/10.1177/019874291203700201

Gage, N. A., Scott, T., Hirn, R., & MacSuga-Gage, A. S. (2018). The relationship between teachers' implementation of classroom management practices and student behavior in elementary school. *Behavioral Disorders*, *43*(2), 302–315. https://doi.org/10.1177/0198742917714809

Konold, T., Cornell, D., Jia, Y., & Malone, M. (2018). School climate, student engagement, and academic achievement: A latent variable, multilevel multi-informant examination. *AERA Open*, *4*(4), 1–17.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What Works Clearinghouse*. https://files.eric.ed.gov/fulltext/ED510743.pdf

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35 year odyssey. *American Psychologist*, *57*(9), 705–717. http://dx.doi.org/10.1037/0003-066X.57.9.705

McLeskey, J., Barringer, M.-D., Billingsley, B., Brownell, M., Jackson, D., Kennedy, M., . . . Ziegler, D. (2017). *High leverage practices in special education*. Council for Exceptional Children.

Moeyaert, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, *52*(2), 191–211. https://doi.org/10.1016/j.jsp.2013.11.003

Mooney, P., Ryan, J. B., Uhing, B. M., Reid, R., & Epstein, M. H. (2005). A review of self-management interventions targeting academic outcomes for students with emotional and behavioral disorders. *Journal of Behavioral Education*, *14*, 203–331. https://doi.org/10.1007/s10864-005-6298-1

Moore, D., Prebble, S., Robertson, J., Waetford, R., & Anderson, A. (2010). Self-recording with goal setting: A self-management programme for the classroom. *Educational Psychology*, *21*(3), 255–265. https://doi.org/10.1080/01443410123466

Peterson, L. D., Young, K. R., Salzberg, C. L., West, R. P., & Hill, M. (2006). Using self-management procedures to improve classroom social skills in multiple general education settings. *Education and Treatment of Children*, *29*(1), 1–21.

Plot Digitizer. (2015). http://plotdigitizer.sourceforge.net

Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology*, *68*, 99–112. https://doi.org/10.1016/j.jsp.2018.02.003

Pustejovsky, J. E., & Swan, D. M. (2018). *Single-case effect size calculator* (Version 0.5) [Web application]. https://jepusto.shinyapps.io/SCD-effect-sizes/

Shadish, W. R., Brasil, I. C., Illingworth, D. A., White, K. D., Galindo, R., Nagler, E. D., & Rindskopf, D. M. (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. *Behavior Research Methods*, *41*(1), 177–183. https://doi.org/10.3758/BRM.41.1.177

Sheffield, K., & Waller, R. J. (2010). A review of single-case studies utilizing self-monitoring interventions to reduce problem classroom behaviors. *Beyond Behavior*, *19*(2), 7–13. https://doi.org/10.17988/BD-13-45.1

U.S. Department of Education Office for Civil Rights. (2014). *Civil rights data collection data snapshot: School discipline*. U.S. Department of Education. https://ocrdata.ed.gov/downloads/crdc-school-discipline-snapshot.pdf

Van Camp, A. M., Wehby, J. H., Martin, B. L. N., Wright, J. R., & Sutherland, K. S. (2020). Increasing opportunities to respond to intensify academic and behavioral interventions: A meta-analysis. *School Psychology Review*, *49*(1), 31–46. https://doi:10.1080/2372966X.2020.1717369

Wakschlag, L. S., Perlman, S. B., Blair, R. J., Leibenluft, E., Briggs-Gowan, M. J., & Pine, D. S. (2017). The neurodevelopmental basis of early childhood disruptive behavior: Irritable and callous phenotypes as exemplars. *American Journal of Psychiatry*, *175*(2), 114–130. https://doi.org/10.1176/appi.ajp.2017.17010045

Wills, H. P., & Mason, B. A. (2014). Implementation of a self-monitoring application to improve on-task behavior: A high-school pilot study. *Journal of Behavioral Education*, *23*(4), 421–434. https://doi.org/10.1007/s10864-014-9204-x

Zimmerman, K. N., Pustejovsky, J. E., Ledford, J. R., Barton, E. E., Severini, K. E., & Lloyd, B. P. (2018). Single-case synthesis tools II: Comparing quantitative outcome measures. *Research in Developmental Disabilities*, *79*, 65–76. https://doi.org/10.1016/j.ridd.2018.02.001